Author’s response to reviews

Title: Prediction models for the risk of gestational diabetes: a systematic review

Authors:

Marije Lamain-de Ruiter (m.deruiter-7@umcutrecht.nl)

Anneke Kwee (a.kwee@umcutrecht.nl)

Christiana A Naaktgeboren (c.naaktgeboren@umcutrecht.nl)

Arie Franx (a.franx-2@umcutrecht.nl)

Karel G M Moons (k.g.m.moons@umcutrecht.nl)

Maria P H Koster (m.p.h.koster@erasmusmc.nl)

Version: 1 Date: 11 Oct 2016

Author’s response to reviews:

Dear Dr. Kengne, We very much appreciate your considerations on our manuscript ‘Prediction models for the risk of gestational diabetes: a systematic review’. We acknowledge the considerations raised by the reviewers and have improved these points. Below is a detailed point-to-point response to the reviewers’ and editorial comments (in blue). Yours sincerely, on behalf of all authors, Marije Lamain – de Ruiter

Reviewer #1. This is a systematic review on prediction models for GDM designed and written well. We thank the reviewer for this compliment.

Reviewer #2. This is a challenging review and the authors have recognised the major limitations - that the samples where these predictive models are derived have varying epidemiological characteristics and risk factors (which is why international risk factor screening varies so much!), and that the diagnostic criteria and screening strategies used for GDM are highly variable. A few issues to address – The definition of GDM ‘any form of hyperglycaemia during pregnancy’ needs to be updated (see current ADA guidelines) We thank the reviewer for this suggestion. The definition is now updated and the reference adjusted (page 5, line 64, reference 1). - They state incidence of GDM has risen rapidly and ranges 3-35% - suspect these are prevalence estimates as the incidence has not risen that significantly. That is correct, we meant prevalence estimates. We have changed the wording accordingly (page 5, line 66). Recently new research was published regarding the prevalence rates, which was added as a new reference [3].

Results: study selection paragraph - number of studies excluded does not add up correctly We apologize for this mistake, the correct numbers are in the figure. We have changed the number in the main text (page 8, line 152).

- What was methodology used to assess risk of bias? Please provide a reference for this method For the risk of bias assessment we have used a method that have been used previously by Smit et al. 2015. We have added the reference (page 8, line 139, reference 18). – Some minor grammatical issues throughout the paper that need addressing We have thoroughly checked and revised our manuscript to improve general readability. This form of systematic review has a lot of short-
comings as you are combining models used for different purposes and this needs to be taken into account eg. some models examine only nulliparous women as these are the women where risk prediction is most challenging, some include porous women; are these models being used to stratify risk of GDM eg. low risk versus high risk and then screening only high risk women at 24-28 weeks, or are they performing universal screening. Do you have any way to account for this within your review? We have not specifically reviewed models for nulliparous versus multiparous women, as the external validation by Lovati et al. 2013 is the only study that exclusively included nulliparous women. All other models developed models for both nulli- and multiparous women. Prediction models for GDM are currently not used in clinical practice. However, we do expect that these models can indeed be used to stratify risk of GDM into low/high risk, and screening only women in the high-risk group for GDM by OGTT at 24-28 weeks as we have stated in the discussion (page 16, line 318). Almost all development studies included in the systematic review used a universal screening approach, except for one (Savvidou et al.) as we have described in the results section (page 11, line 197). As these two studies were the exceptions, we have not taken this into special account in our systematic review. Variable screening and diagnostic criteria is a significant limitation - can you group results and do a sub-analysis according to criteria used? The studies included in our review used eight different diagnostic outcome criteria and/or variable screening methods. With a total number of 17 studies included, we expect subgroups to be so small that this will severely limit the value of sub-analysis. Therefore, we have chosen not to perform sub-analysis, although we do agree with the reviewer that this would be interesting. A comment on this was added in the discussion section (page 15, line 285). Studies examining predictive models also use different methods for their sensitivity and specificity or probability cut offs. eg. Nanda (model 4) explored fixed false positive rates. Has this been considered where you refer to the "best predictive models" because this makes them difficult to compare. We agree with the reviewer that it is difficult to mutually compare the studies as different methods and cut-offs have been used to report sensitivity and specificity. With the information provided in the original publication it is almost impossible to consider the implications of these differences, therefore we have recommended to perform an external validation and direct comparison of the developed prediction models to make a robust comparison.